Figures and figure supplements

Eco-evolutionary dynamics of adapting pathogens and host immunity

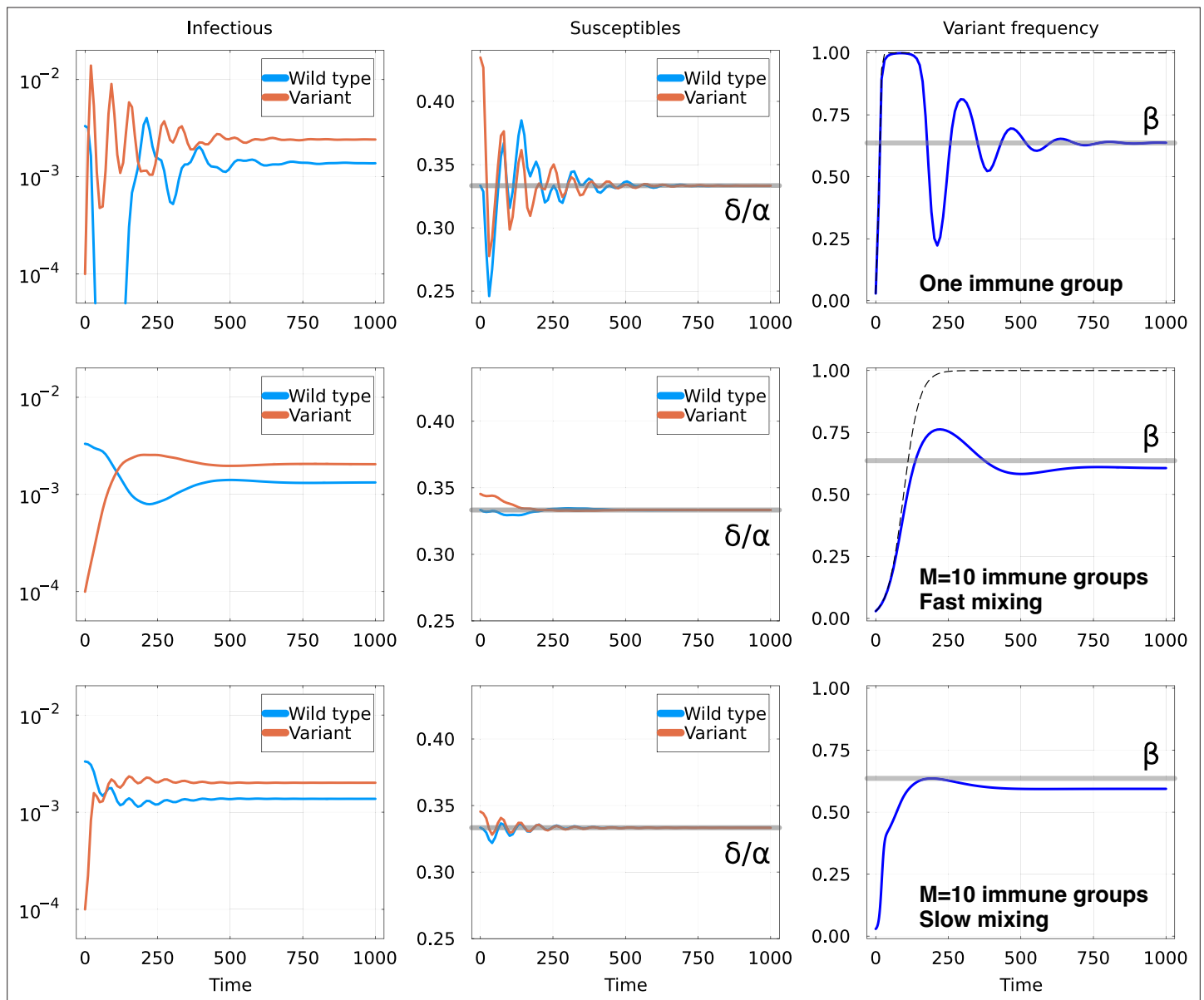**Pierre Barrat-Charlaix and Richard A Neher.**

Figure 1. Invasion of an immune escape mutant. Top row: one immune group, Middle row: $M = 10$ immune groups and fast mixing $C_{ij} = 1/M$ and Bottom row $M = 10$ immune groups and slow mixing $C_{ij} = 1/10M$. Other parameters are the same for all rows: in units of δ, we set $\alpha = 3$ and $\gamma = 5 \cdot 10^{-3}$, and $f = 0.65$, $b = 0.8$, $\varepsilon = 0.01$. For both rows, graphs represent: Left: number of hosts infectious with the wild-type and the variant; Middle: number of hosts susceptible to the wild-type and the variant, with the equilibrium value δ/α as a gray line; Right: fraction of the infections due to the variant. The thick gray line shows the expected equilibrium frequency β in the case with one immune group, given in **Equation 6**. The dashed line shows the trajectory of a constant fitness logistic growth with the same initial growth rate.
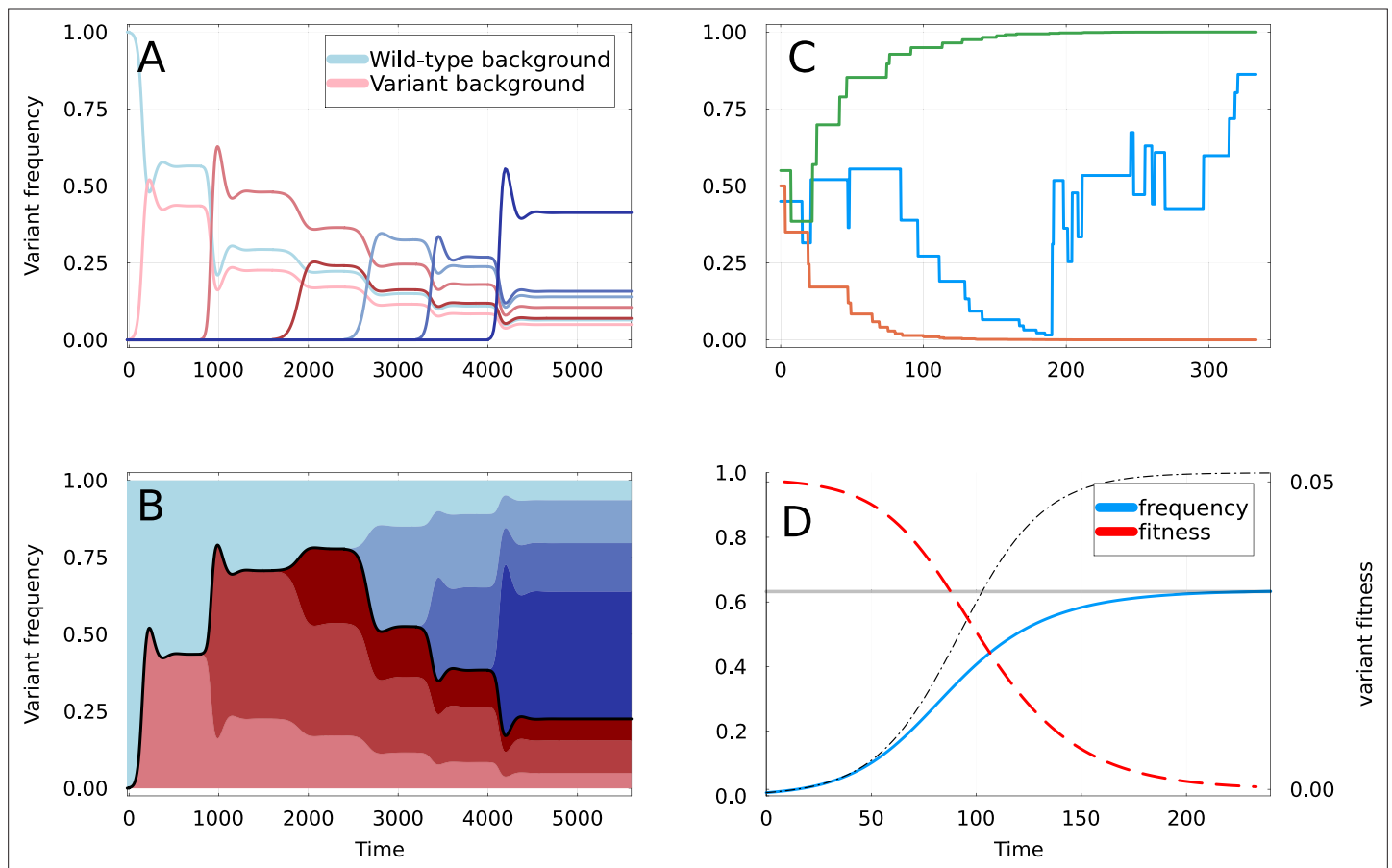
Figure 2. Dynamics of partial sweeps and subsequentfixation. (**A**) Simulation of Susceptible-Infected-Recovered (SIR) *Equation 1* & *Equation 2* with additional strains appearing at regular time intervals. The fraction of infections (frequency) caused by each strain is shown as a function of time. The first strain to appear at $t = 0$ is the variant of interest, and curves are shown in shades of red if they appear on the background of this variant, and of blue if they appear on the background of the wild-type. (**B**) Same as A but with frequencies stacked vertically. The black line delimiting the red and blue areas represents the frequency at which the mutations defining the original variant are found. (**C**) Three realizations of the random walk of *Equation 9*, all starting at $x \simeq 0.5$. Two instances converge rapidly to frequencies 0 and 1, corresponding to apparent selective sweeps, while the remaining one oscillates for a longer time. (**D**) Representation of a partial sweep using the expiring fitness parametrization of *Equation 11*. The frequency x of the variant is shown as a blue line saturating at value β (gray line). The thin dashed line shows a selective sweep with constant fitness advantage s_0. The fitness s is a red dashed line, using the right-axis.
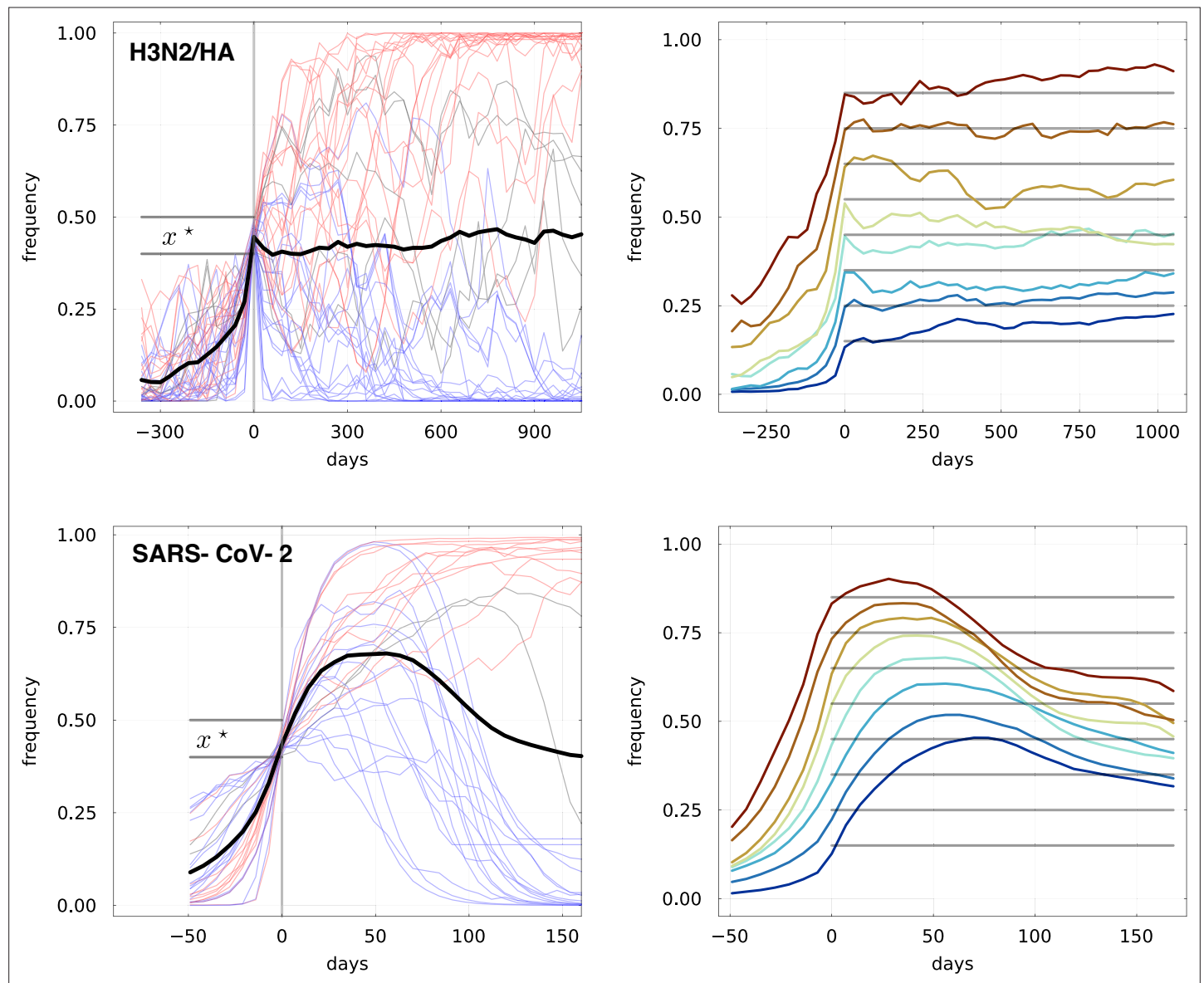
Figure 3. Retrospective analysis of predictability of viral evolution: frequency trajectories of all amino acid substitutions that are observed to rise from frequency 0 to $x^\star$ for Top: influenza virus A/H3N2 from 2000 to 2023, and Bottom: SARS-CoV-2 from 2020 to 2023. Left: all trajectories for $x^\star = 0.4$, with blue ones ultimately vanishing and red ones ultimately fixing. The average of all trajectories is shown as a thick black line. Right: showing only the average trajectories for different values of $x^\star$ (gray lines).
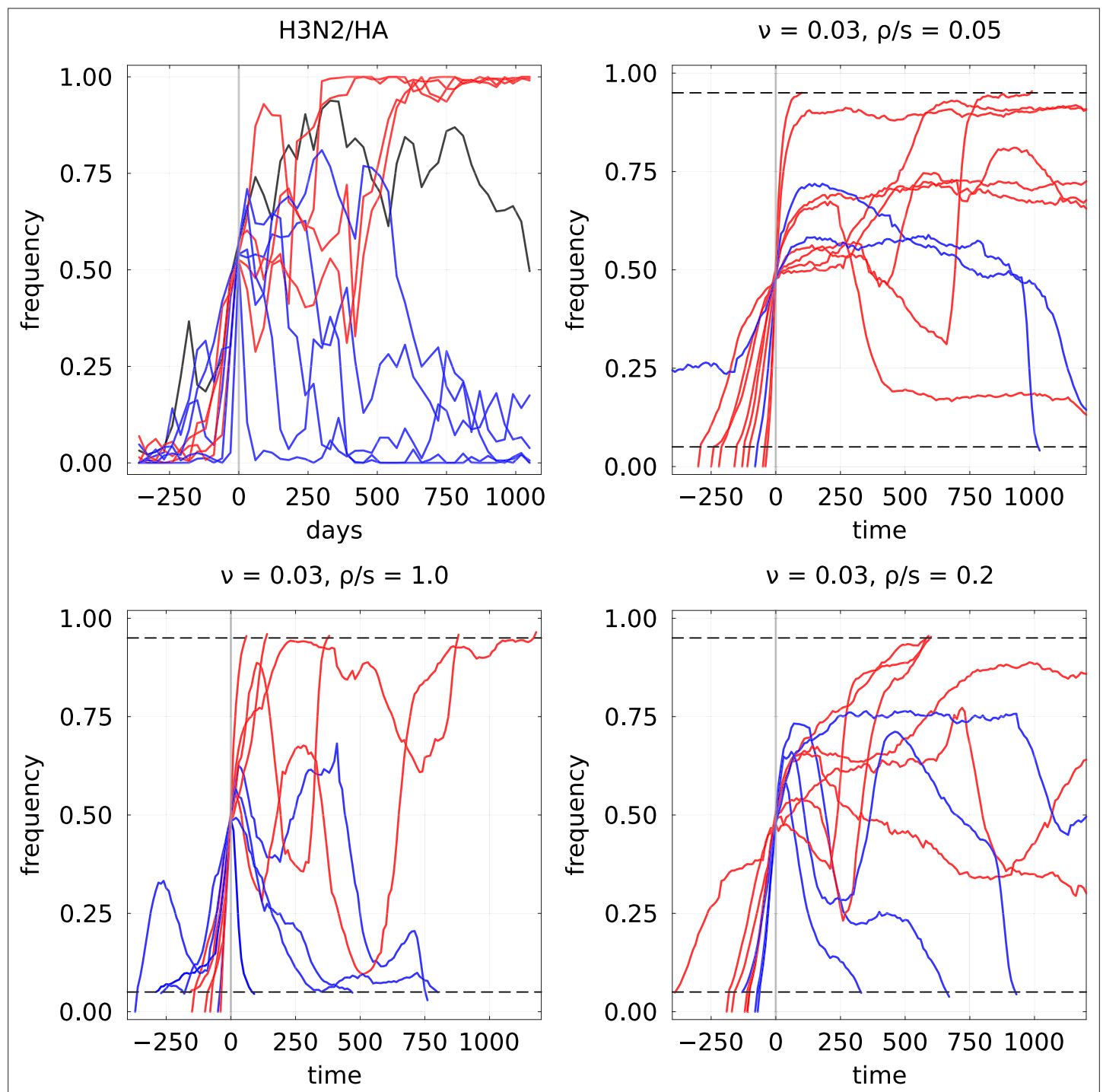
Figure 3—figure supplement 1. Example of mutation frequency trajectories that are increasing up to a frequency of 0.5 for H3N2/HA influenza and the expiring fitness model. For the latter, parameters used are $\alpha = s_0 = 0.03$ and three values of ρ to illustrate different clonal interference regimes. In each case, 10 randomly selected trajectories are plotted, with blue color indicating final loss and red final fixation.
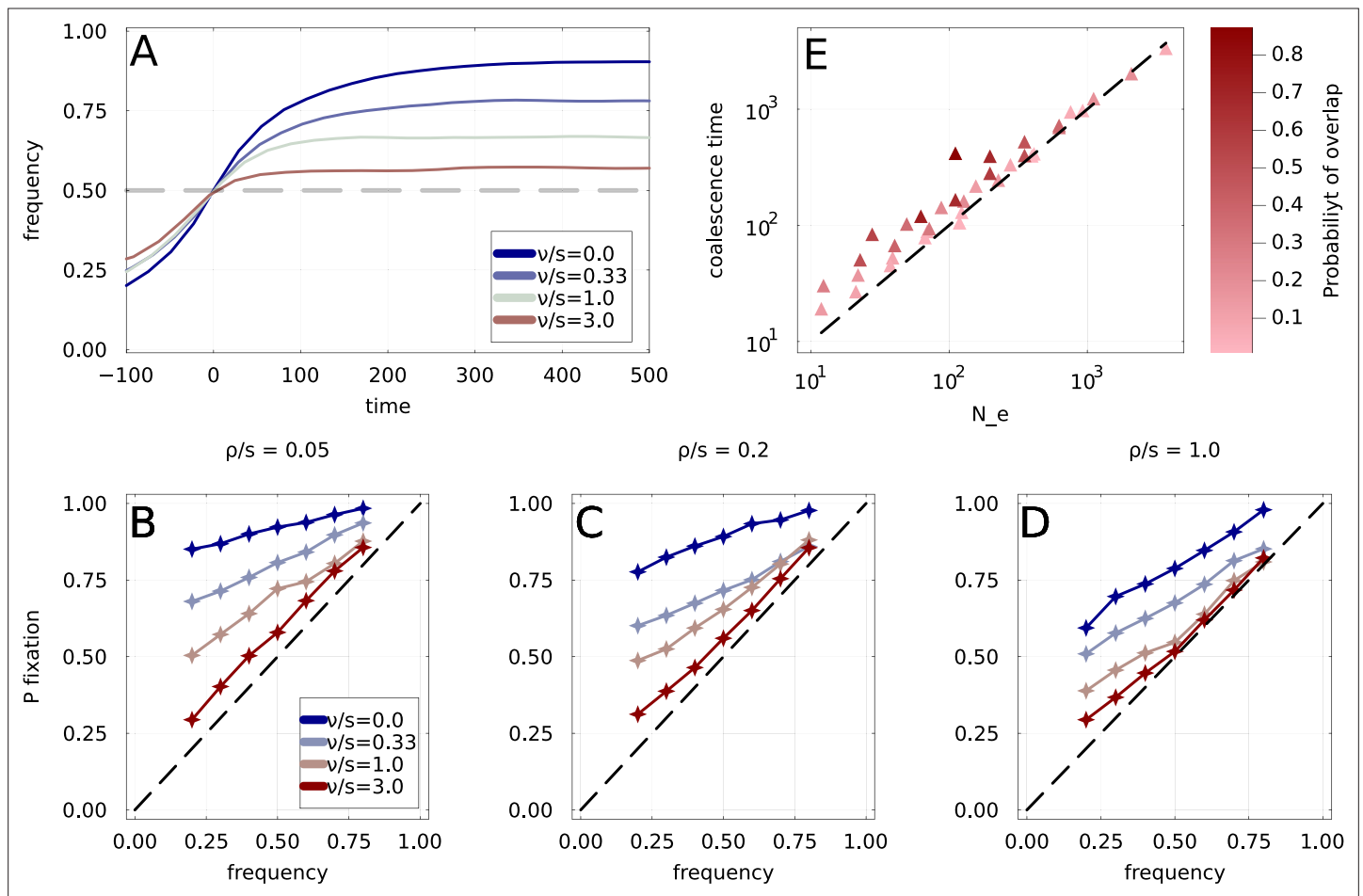
Figure 4. Simulations under the Wright-Fisher model with expiring fitness. (**A**) Average frequency dynamics of immune escape mutations that are found to cross the frequency threshold $x^* = 0.5$, for four different rates of fitness decay. If the growth advantage is lost rapidly (high ν/s_0), the trajectories crossing x^* have little inertia, while stable growth advantage (small ν/s_0) leads to steadily increasing frequencies. (**B, C, D**) Ultimate probability of $p_{fix}(x)$ of trajectories found crossing frequency threshold x. Each panel corresponds to a different rate of emergence of immune escape variants, with four rates of fitness decay per panel. Increased clonal interference ρ/s_0 and fitness decay ν/s_0 both result in a gradual loss of predictability. We use $s_0 = 0.03$. (**E**) Time to most recent common ancestor T_{MRCA} for the simulated population, as a function of the prediction obtained using the random walk $N_e = 1/\rho\langle\beta^2\rangle$. Points correspond to different choices of parameters ρ and P_β, and a darker color indicates a higher probability of overlap as computed in Appendix 2.2.
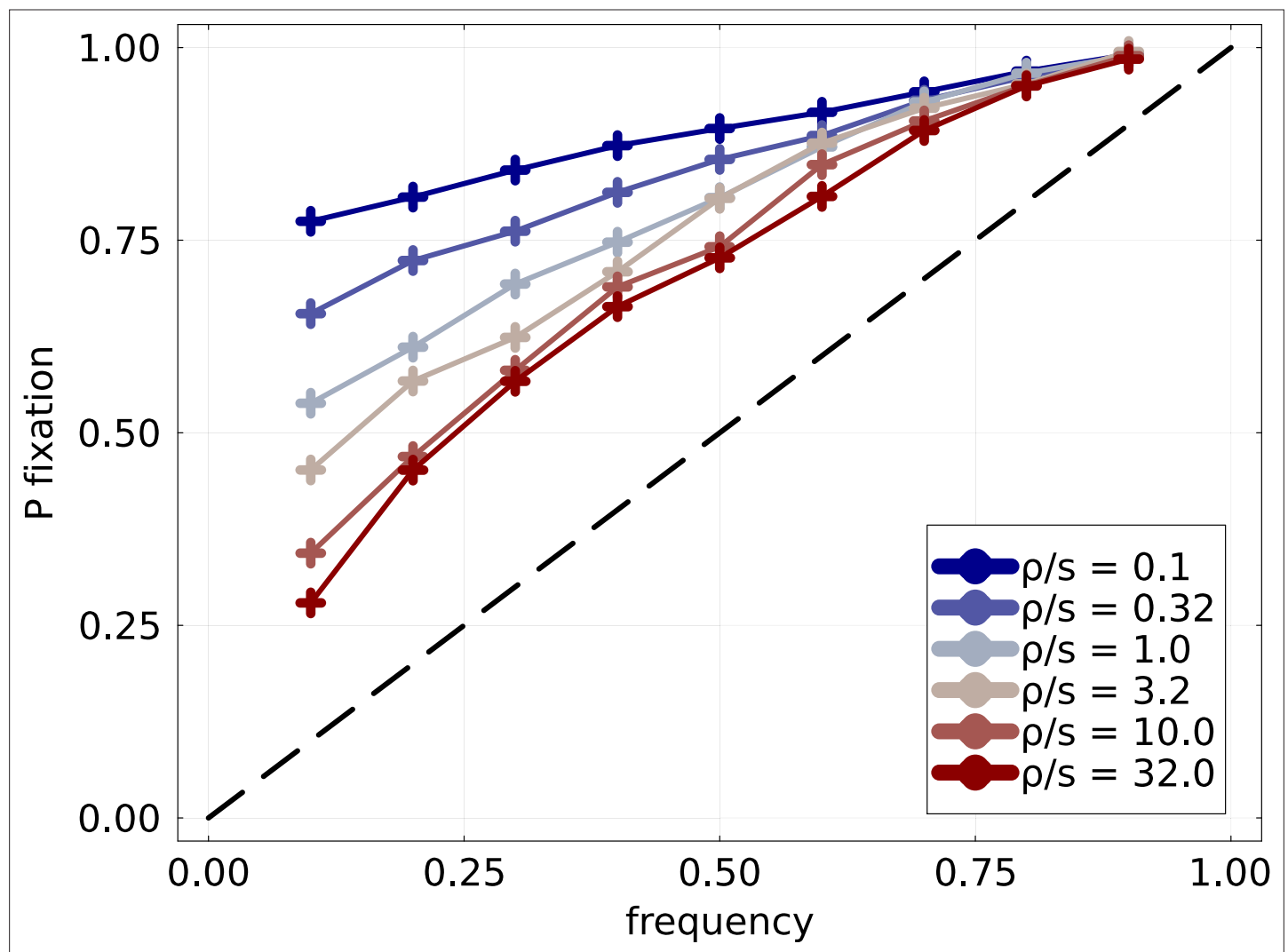
Figure 4—figure supplement 1. Probability of fixation of mutation Probability of fixation of mutations $p_{fix}(x)$ of mutation frequency trajectories found crossing the frequency threshold x. Fitness effects are exponentially distributed with fixed scale $s_0 = 0.03$. The blue to red gradient in colors corresponds to the increasing rate ρ at which mutations are introduced. Strong clonal interference regime is obtained when $\rho/s_0 > 1$, in which case good mutations are introduced in close succession and compete for fixation. At low ρ/s_0, trajectories are very predictable and an increasing trajectory almost certainly fixes. Even for the highest ρ/s_0, $p_{fix}(x)$ remains significantly larger than x and dynamics are visibly not neutral.
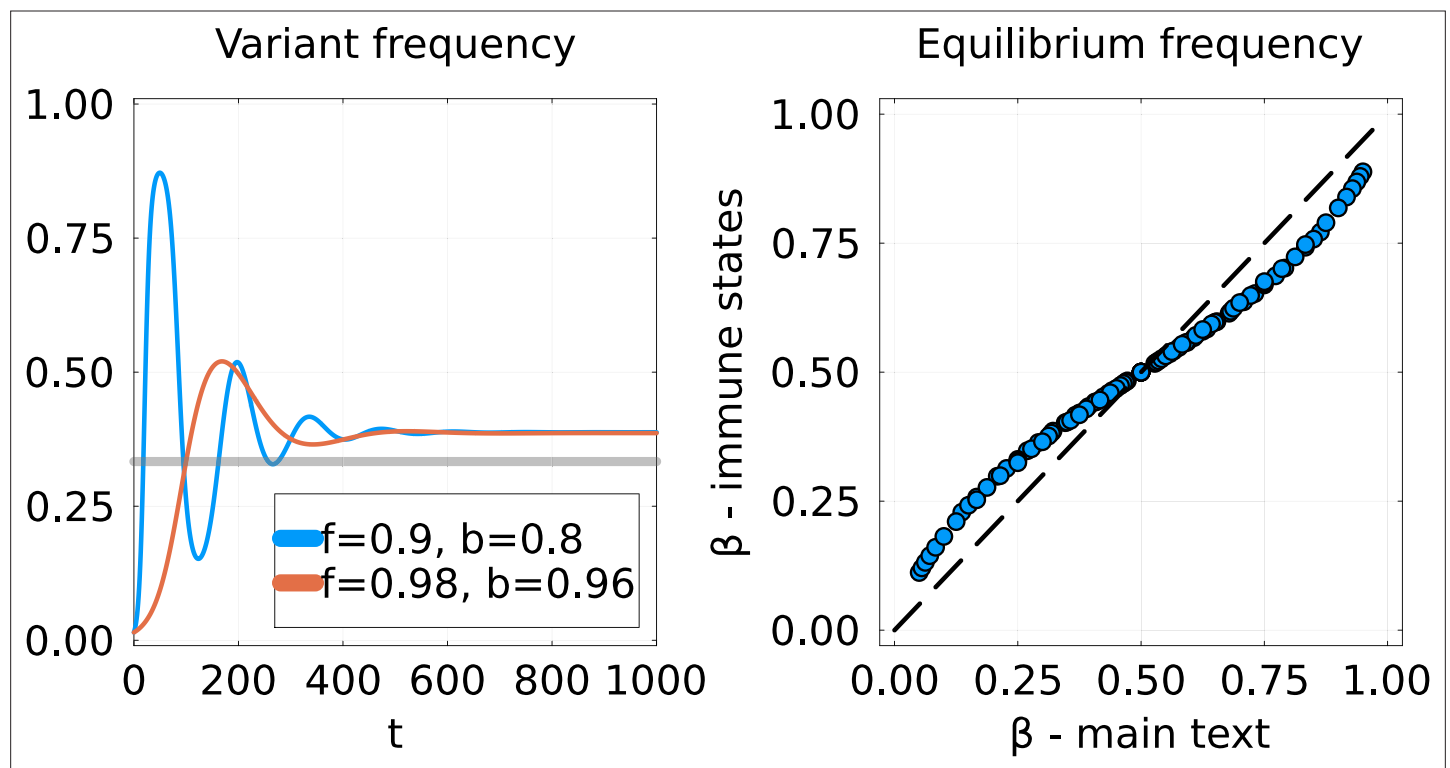
Appendix 1—figure 1. Comparison of SIR model implementations. Left: Dynamics of the frequency of the variant for the Susceptible-Infected-Recovered (SIR) model from *Equations 25; 26* using the invasion scenario from the main text. Two 2×2 cross-immunity matrices are used, with off-diagonal parameters f and b chosen to give the same equilibrium. The gray line represents the equilibrium that would be obtained using the model of the main text. Right: Equilibrium frequency β for this new SIR model (y-axis) versus the β from the main text (x-axis). Each point corresponds to a given pair (f, b).
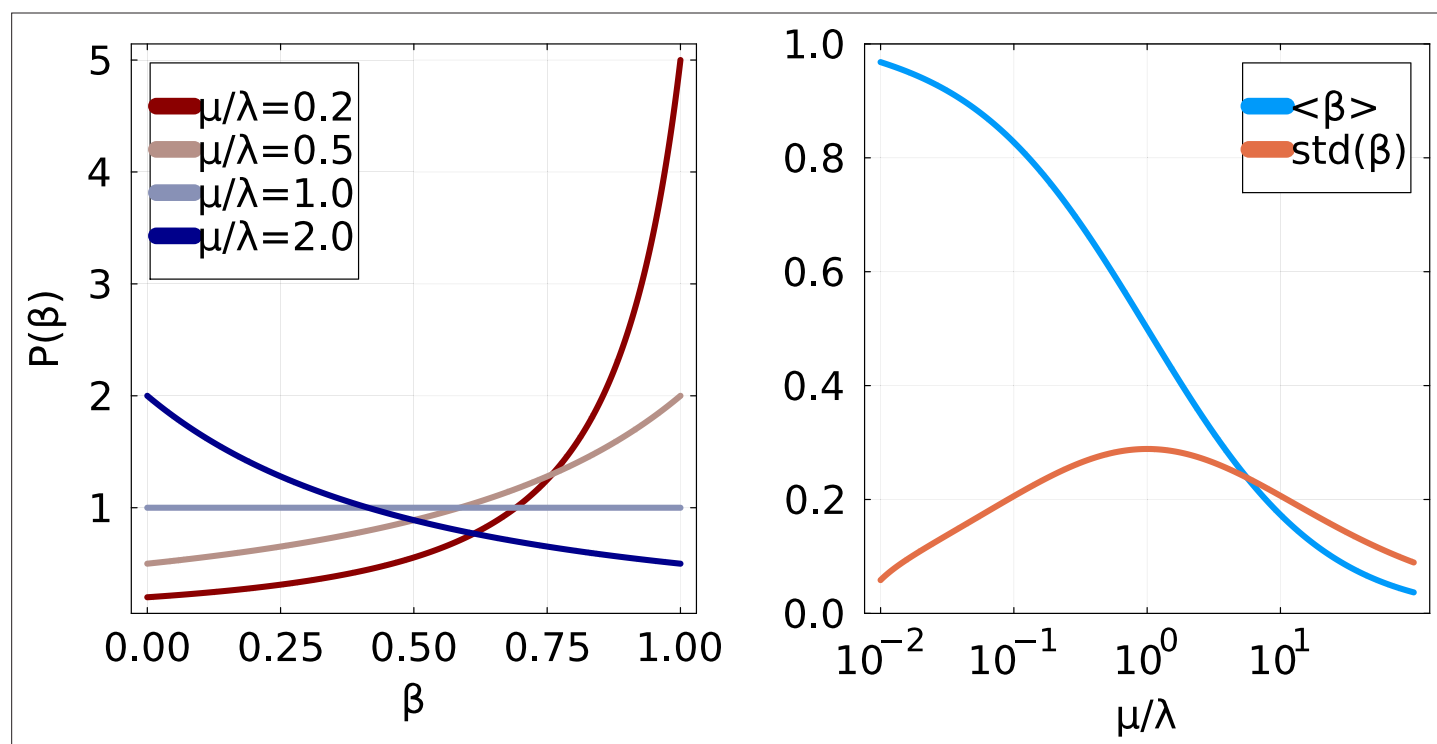
Appendix 1—figure 2. Distribution of partial sweep size β if $1 - f$ and $1 - b$ are exponentially distributed with respective scales μ and λ. Left: Probability distribution function $P(\beta)$ for various values of μ/λ. Right: Mean and standard deviation of β as a function of μ/λ.
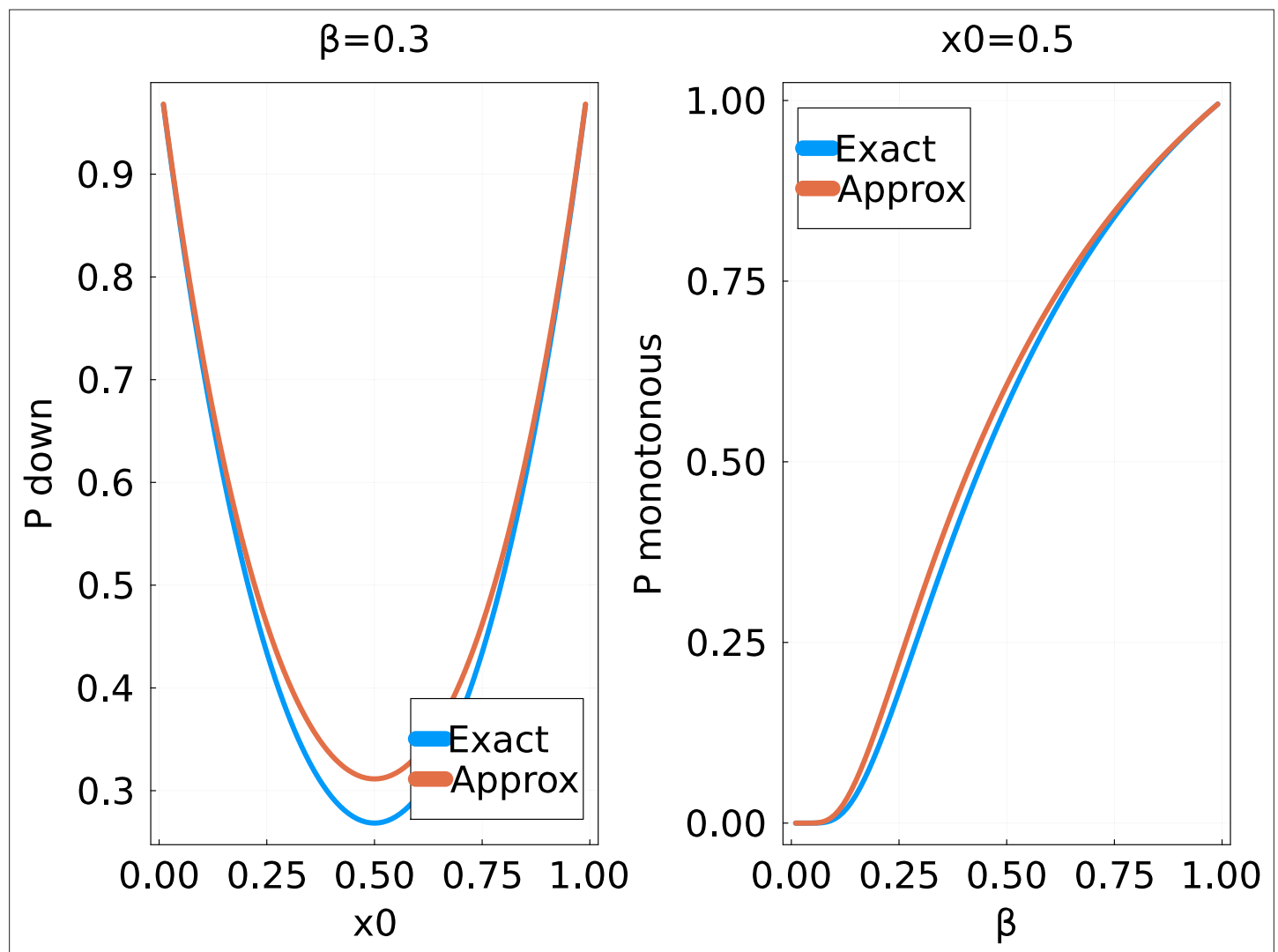
Appendix 2—figure 1. Probability of a strictly monotonous trajectory in the random walk of the main text, as a function of β (fixed) and the initial value x_0. The 'exact' solution is obtained by numerically computing the product in *Equation 41* up to $t = 100$.
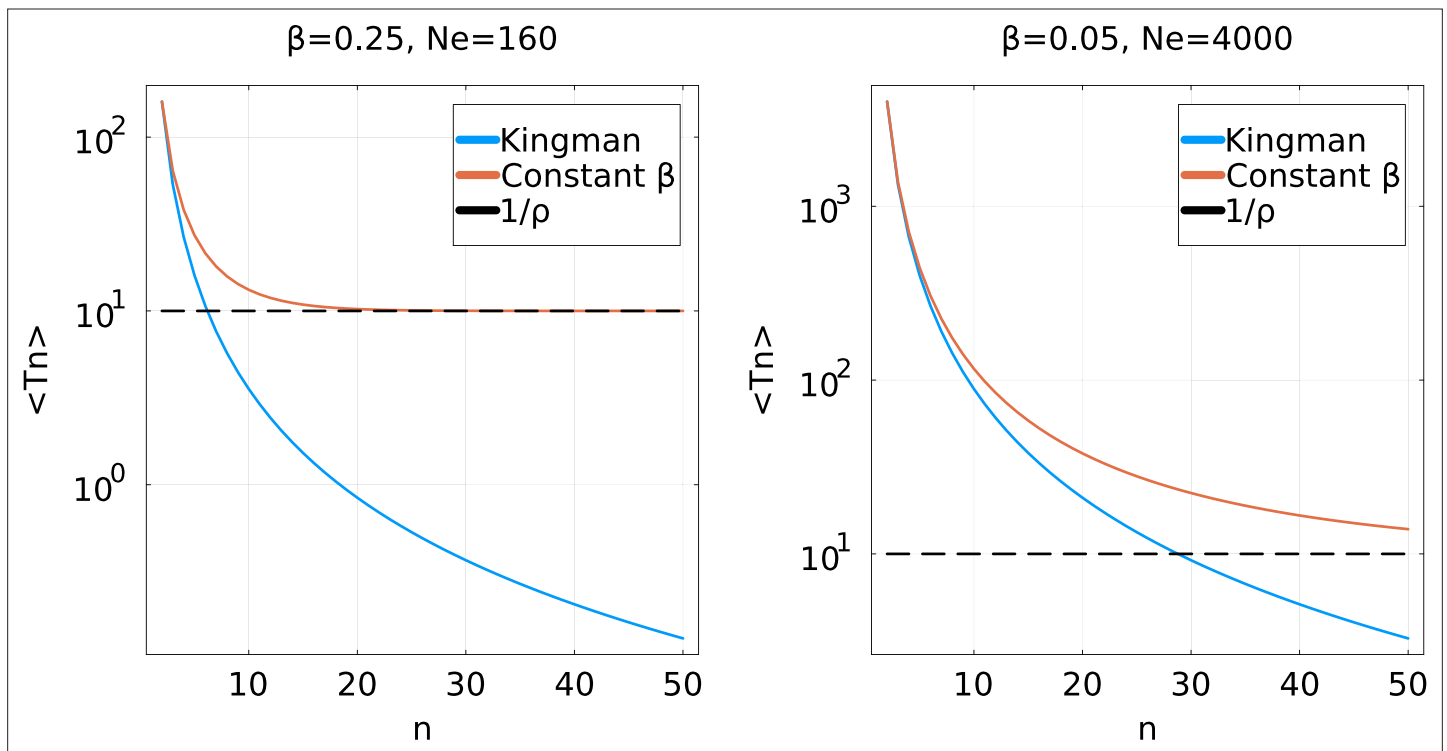
Appendix 2—figure 2. Average coalescence times $\langle T_n \rangle$ for a partial sweep coalescent with effective population size N_e and a Kingman coalescent with population size N_e. For simplicity, a constant β is used: Left: a high value $\beta = 0.25$; Right: a low value $\beta = 0.05$. For low β, the two coalescent processes are very similar until a high n. They considerably differ if β is larger. Note that for the partial sweep process, T_n never goes below ρ^{-1}.
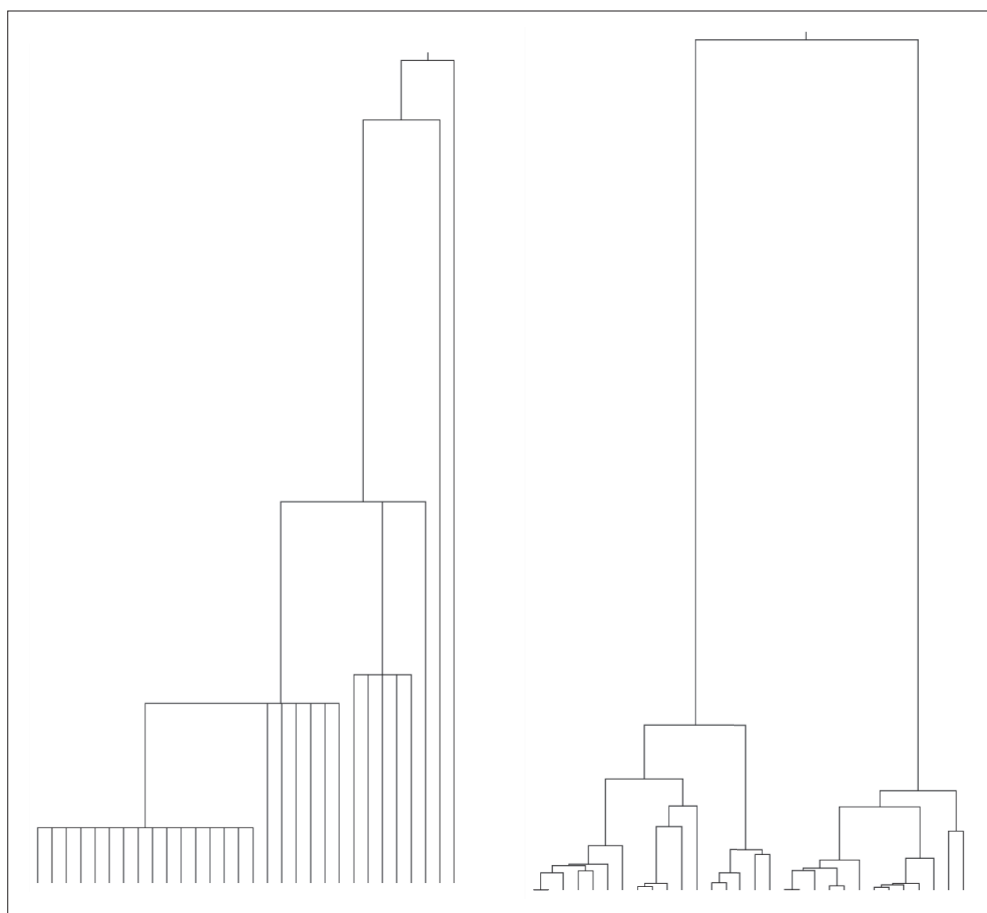
Appendix 2—figure 3. Realizations of different coalescence processes for 30 lineages (leaves). Left: Partial sweep coalescent, with constant $\beta = 0.4$ and $\rho = 0.00625$ such that $N_e = (\rho\beta^2)^{-1} = 1\,000$. Right: Kingman coalescent with population size $N = N_e = 1\,000$.